

DOCUMENT RESUME

ED 086 243

IR 000 097

AUTHOR Kowitz, Gerald T.; And Others
TITLE From ERIC Source Documents to Abstracts: A Problem in Readability.
PUB DATE 29 Nov 73
NOTE 8p.; Paper presented at the Rocky Mountain Educational Research Association (Tucson, Arizona, November 29, 1973)
EDRS PRICE MF-\$0.65 HC-\$3.29
DESCRIPTORS *Abstracts; Clearinghouses; Information Sources; Information Theory; *Readability; Reading Comprehension; Sampling
IDENTIFIERS *ERIC; Flesch Readability Index

ABSTRACT

The use of readability principles in the ERIC guidelines for abstracting suggests that abstracts should be more readable than source documents. However, information theory suggests that the compression of information would result in less readable abstracts. Abstracts and their source documents were selected randomly from each of four clearinghouses also selected randomly. Sample size was determined by a calculation of power. The Flesch RE formula was used to calculate readability scores. Results of an analysis of variance indicated that abstracts were less readable than source documents but their mean grade levels of 11.7 to 12.7 were within the expected reading ability level of users of the ERIC system. (Author)

ED 086243

From ERIC Source Documents to Abstracts:
A Problem in Readability*

Gerald T. Kowitz
Gladys B. Dronberger
William H. Graves

College of Education
University of Oklahoma

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

Some years ago it was popular to discuss an "explosion of knowledge." Dramatic figures were given concerning the rate of publications in scientific and technical fields. (Levine, Worboys and Taylor, 1973). Some critics have claimed that much was replication or repetition of older work and that the addition of new materials was not great. However, there can be no doubt that the advance of scientific and technological knowledge has produced an increase in the volume of materials. The student, as well as the professional researcher, is faced with a major problem in identifying papers pertinent to his field. Scientific, technical and scholarly societies traditionally have approached the problem by providing abstracts of publications in their particular fields. Some of these abstracting services have functioned for nearly a hundred years. Probably, the earliest was established by the Abbott Laboratories in 1888. The Chemical Abstract Service was started in 1907 and the Psychological Abstracts in 1927 (Kruzas, 1971). There are now indexes to abstracting services, which in themselves, have produced a formidable task for the scholar who would want to acquaint himself with a particular topic.

*Presented at Rocky Mountain Educational Research Association, Tucson, Arizona, November 29, 1973.

IR 000 097

Most scholarly societies have relied upon volunteers from their membership to write abstracts of journal articles, monographs, and pertinent resources. The general pattern has been for the society to provide guidelines which encouraged writers to be brief, to avoid technical terms, and to make the abstracts readable. The purpose of making them readable is to assist the person who is not an expert in the field in identifying articles which will be of the greatest use. More readable abstracts also are valuable to the researcher who must process rapidly a great body of information to select pertinent documents. Abstracts have become such a part of academic life that few people question their usability. In fact, it is generally assumed that they are functioning well.

The advent of the computer and of programs which could retrieve and classify abstracts for the scholar have posed a number of interesting problems. Solutions have been found to the library problem of indexing, and classifying abstracts. A barrier still exists between the scholar and the literature if abstracts, once they are located, are not precise and readable. An interesting conflict exists. One dimension of the conflict can be deduced from the field of readability.

By preparing brief abstracts of a highly readable nature, the size of the audience should be increased. Thus, the information contained in the literature of an academic or scientific field should be made more useful to a larger audience. Second, the problem of abstracting can be viewed from the position of information theory. The writing of short abstracts reduces the redundancy normally found in language; it increases the density of information and the rate with which it will be transmitted. These factors should make the abstract materials much less

useful to those not thoroughly trained in the field.

In surveying several computer-based retrieval services, it becomes apparent that while they function well, they have not approached their goal. Most services have experienced low usage for computer based searches even though they can identify relatively large numbers of articles in a brief period of time. Some have suggested that the reason is the common aversion in many fields to the use of the computer. Others emphasize the fact that a computer search normally identifies a large number of abstracts. The common excuse is that upon receiving a stack of computer printouts the user plans to look at them later--only mañana never comes; the problem of the ability of the abstracts to transmit information is obviated. Perhaps, if they were highly readable, they could be scanned rapidly and not laid aside. Thus, the readability of abstracts is a beginning point for examining the usability of a system.

Beginning about 1966 the U. S. Office of Education began to develop a large data base of materials relevant to the full range of educational endeavors. Early in the planning two decisions were made. One was to provide abstracts to permit a rapid search through the many documents; the second was to have the abstracts put on computer tape for rapid retrieval. At that time, no workable programs for abstract searches by computer existed. Indeed, many difficulties were experienced in getting operational, computer-based retrieval systems. However, in 1967, the University of Oklahoma, using its General Information Processing System (GIPSY), had all existing ERIC abstracts available for computer based searches. The functioning of this system has been described elsewhere in the literature, (Kowitz, 1970; Kowitz, et al, 1971; Kowitz, et al, 1973).

Readability

Many formulas have been developed to implement the study of readability. Most began with the idea that if the readability level of material can be matched reasonably well with the skills of the reader he will profit more from them. The idea has been used both in the teaching of reading and in the development of materials for teaching in other fields. Some have objected to the use of readability formulas to write materials because they produce poor literature (Bormuth, 1968). Others have suggested that while the formulas should not be used directly in writing, knowledge of the elements of the formula can be used in developing guidelines which will produce more readable materials (Klare, 1963).

ERIC

The ERIC system presented a second interesting phase of the problem. The abstracts were written by many people ranging in occupation from housewives to graduate students. Clearinghouses were established in specific areas of education. Each of them has the assignment of collating documents within its topic and writing abstracts of them. The central office of ERIC provided guidelines for writing abstracts to be used by all clearinghouses.

Documents, Abstracts and Readability

The present study identified several clearinghouses on a random basis within certain constraints. Only those clearinghouses which had a long history were to be used. Thus, it would be possible to gauge the consistency of the readability of the abstracts. A list of all abstracts from the selected clearinghouses was assembled and sample abstracts were drawn at random. Constraints also were applied

to the selection of abstracts. No abstract would be used if it was primarily a listing or bibliography, contained less than one hundred words, or referred to a curriculum guide or other catalog materials. Subsequently, the source documents, available on microfiche, were sampled. The Flesch Reading Ease formula was then applied to the body of the abstract and to the sample pages of the source document.

Because the number of documents in ERIC-RIE is so large, a problem existed in sampling. The problem was approached through the concept of estimation of the power of the analysis of variance. In an earlier study, (Kowitz, et al, 1973) the range of readability of abstracts from a clearinghouse was found to be approximately six grade levels. Therefore, it was concluded that a difference of 1.50 standard deviations would indicate significant differences at the 5% level of confidence. Given this information, it was determined that equal sized samples of 12 abstracts drawn at random from each clearinghouse would provide an adequate base for statistical inference. Similarly, when considering the list of clearinghouses with the longest histories, it was determined that four clearinghouses would provide an adequate base for this first investigation. The clearinghouses identified were:

1. Adult Education
2. Educational Management
3. Educational Media and Technology
4. Rural Education and Small Schools

The reference numbers of the abstracts listed by clearinghouses are available on request. The Flesch Readability Index was applied to both source document samples and to their abstracts.

The following hypotheses were tested: 1

1. Abstracts should have a lower reading level than the original documents.
2. The reading level of material processed by a clearinghouse from one area of education, is different from that processed from other areas.
3. The clearinghouses that produced abstracts with the lower reading levels processed documents of the lowest reading levels.
4. Different clearinghouses produced abstracts with different readability levels.
5. There is a difference among the reading levels of the documents abstracted by different clearinghouses.
6. There is a difference in the variability of the readability of abstracts produced by different clearinghouses.

Results

The means and standard deviations for source documents and abstracts are presented in Table 1. The results of the correlated measures analysis of variance are in Table 2 and the post hoc pair-wise comparison of means in Table 3. (Note: Tables not included in Paper)

Because of the conflicts described earlier between theories of readability and information, a non-directional test was used. Results of testing hypotheses indicated that the abstracts had a more difficult level of readability than did the samples from the original documents. All differences between the readability of abstracts and source documents, hypothesis 1, were significant except for the Clearinghouse for Adult Education. The test of hypothesis 2, the reading levels of

materials produced in the different areas, failed to reveal a significant difference. Hypothesis 3, which suggested that the abstracts with the lowest reading levels would result from documents with the lowest reading levels was not supported by the analysis. There was a significant interaction between clearinghouses and the production of abstracts. This is illustrated in Figure 1.

The readability of abstracts produced by different clearinghouses hypothesis 4, resulted in one statistically significant difference, between the abstracts from Educational Management and Adult Education.

The source documents abstracted by different clearinghouses, hypothesis 5, failed to show significant differences in their reading levels.

The last hypothesis was tested using the F-maximum test for homogeneity of variances among the readability scores of abstracts and source documents from the different clearinghouses. The results were not statistically significant.

Conclusion

The use of guidelines for preparing abstracts appears to be an effective tool. However, the general assumption that abstracts provide a more readable as well as a briefer statement of the original document, was not supported. In general, however, the grade level equivalents indicated that both the abstracts and the source documents were in the desirable range of readability. It would seem, therefore, that the ERIC-RIE system should be a valuable source for educators in locating documents in which they are interested.

References

- Bormuth, John R. "New measures of grammatical complexity," in Goodman, K. S. (ed.), *The Psycholinguistic nature of the reading process*. Detroit: Wayne University Press, 1968.
- Klare, G. R. *The measurement of readability*. Ames, Iowa: University of Iowa Press, 1963.
- Kowitz, G. T. *Information as technology for education*, SRIS, 1970, Winter, 6-9.
- Kowitz, G. T., Reeves, R. R., and Prather, J. R. *The GIPSY/ERIC retrieval system*. Monograph V. *Information Science Series*, 1971, EDRS No. ED047-763.
- Kowitz, G. T., Reeves, R. R., Dronberger, G. B., Bishop, N. E., and Abbott, N. T. *Values in guidance: Can ERIC help? Measurement and Evaluation in Guidance*, 1973, 6, 47-53.
- Kruzas, A. T. *Encyclopedia of information systems and services*, Ann Arbor: Edwards Brothers, 1971.
- Levine, N., Worboys, C., and Taylor, M. *Psychology and the 'Psychology' textbook: A social demographic study*. *Human Relations*, 26, 4, 467-478, 1973.